# 4. Convergence en loi

La convergence en loi est une notion de convergence pour les variables aléatoires qui est plus faible que la convergence presque sûre, et qui capture l'idée suivante : $(X_n)_{n\in\mathbb{N}}$ *converge en loi* (ou *en distribution*) vers une variable $X$ si les probabilités relatives à $X_n$ (mais pas les valeurs de $X_n$) convergent vers les probabilités relatives à $X$. Pour des raisons topologiques, il faut faire un peu attention à la définition précise de convergence des probabilités. Les conditions suivantes s'avèrent être équivalentes pour une suite de variables aléatoires réelles $(X_n)_{n\in\mathbb{N}}$ :

(1) *Convergence des observables.* Pour toute fonction continue bornée $f : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[f(X_n)] \to_{n\to\infty} \mathbb{E}[f(X)].$$

(2) *Convergence des probabilités.* Pour toute partie mesurable $A$ telle que $\mathbb{P}[X \in \partial A] = 0$,

$$\mathbb{P}[X_n \in A] \to_{n\to\infty} \mathbb{P}[X \in A].$$

(3) *Convergence des fonctions de répartition.* Pour tout point $t$ tel que $F_X$ soit continue en $t$,

$$F_{X_n}(t) \to_{n\to\infty} F_X(t).$$

(4) *Convergence des transformées de Fourier.* Pour tout réel $\xi$,

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\xi X_n}] \to_{n\to\infty} \mathbb{E}[\mathrm{e}^{\mathrm{i}\xi X}].$$

On admettra l'équivalence de ces conditions, et on dira si elles sont vérifiées que $(X_n)_{n\in\mathbb{N}}$ converge en loi vers $X$ (notation : $X_n \rightharpoonup_{n\to\infty} X$). Notons que si $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires qui converge presque sûrement vers $X$, alors par le théorème de convergence dominée, on a convergence des observables : la convergence presque sûre est donc plus forte que la convergence en loi. Si les variables $X_n$ et $X$ sont à valeurs dans $\mathbb{Z}$, alors la convergence en loi est simplement donnée par :

(5) *Convergence locale.* Pour tout $k \in \mathbb{Z}$,

$$\mathbb{P}[X_n = k] \to_{n\to\infty} \mathbb{P}[X = k].$$

Si $X \sim \mu$, on notera aussi la convergence en loi $X_n \rightharpoonup_{n\to\infty} \mu$.

### 1. Convergence vers la loi de Poisson.

On fixe un paramètre $\lambda > 0$ et on travaille avec des entiers $n \geq \lambda$.

(1) Dresser les histogrammes empiriques de $N$ tirages $X_1, \ldots, X_N$ d'une loi binomiale $\mathrm{Bin}(n, \frac{\lambda}{n})$ avec $N = 1000$, $n = 50, 1000$ et $\lambda = 1, 3, 5$. Comparer avec les histogrammes des lois de Poisson $\mathrm{Poi}(\lambda)$.

(2) Montrer avec le critère de convergence locale qu'on a effectivement $\mathrm{Bin}(n, \frac{\lambda}{n}) \rightharpoonup_{n\to\infty} \mathrm{Poi}(\lambda)$.

(3) Donner une autre preuve reposant sur la convergence des transformées de Fourier.

**2. Théorème central limite.**

Le *théorème central limite* est un énoncé de convergence en loi pour la différence renormalisée entre une moyenne empirique de variables indépendantes et de même loi, et la moyenne théorique. Ainsi, si $X_1, X_2, \ldots$ sont des variables indépendantes dont la loi commune $\mu$ admet un moment d'ordre 2 :

$$\int_{\mathbb{R}} x^2\,\mu(\mathrm{d}x) < +\infty,$$

et si $M_n = \frac{X_1+X_2+\cdots+X_n}{n}$ et $m = \mathbb{E}[X_1]$, alors

$$\sqrt{n}\left(\frac{M_n - m}{\sigma}\right) \rightharpoonup_{n\to\infty} \mathrm{N}(0,1),$$

où

$$\sigma^2 = \left(\int_{\mathbb{R}} x^2\,\mu(\mathrm{d}x)\right) - \left(\int_{\mathbb{R}} x\,\mu(\mathrm{d}x)\right)^2$$

désigne la variance de la loi $\mu$, et $\mathrm{N}(0,1)$ est la loi gaussienne de densité $f(x) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}$ sur $\mathbb{R}$. La preuve repose sur le critère des transformées de Fourier, et sur le développement de Taylor $\mathbb{E}[\mathrm{e}^{\mathrm{i}\xi(X_1-m)}] = \exp\!\left(-\frac{\sigma^2\xi^2}{2} + o(1)\right)$.

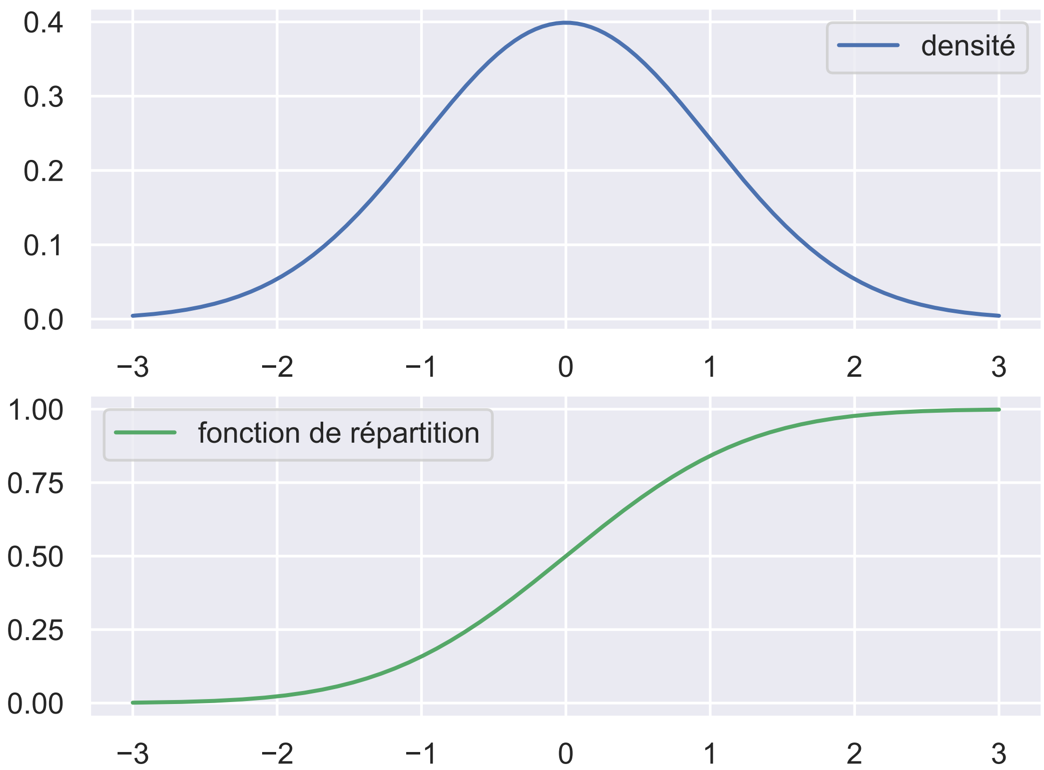


Fig. 4.1. Densité et fonction de répartition de la loi normale $\mathrm{N}(0,1)$.

(1) Écrire un programme qui tire un échantillon $(X_1, \ldots, X_N)$ de variables aléatoires de même loi, et qui calcule le vecteur formée par toutes les moyennes recentrées et renormalisées :

$$\left(Z_n = \sqrt{n}\,\left(\frac{\frac{X_1+\cdots+X_n}{n} - m}{\sigma}\right)\right)_{n\in[\![1,N]\!]}.$$

On pourra faire des essais avec différentes lois : par exemple, la loi uniforme sur $[0,1]$, la loi exponentielle $\mathrm{Exp}(1)$, et la loi de Bernoulli $\mathrm{Ber}(\frac{1}{2})$.

(2) Tracer pour la loi choisie le graphe de $n \mapsto Z_n$ pour $n \in [\![1, 5000]\!]$. A-t-on une convergence presque sûre ?

(3) Tracer la fonction de répartition empirique d'un échantillon de $M = 1000$ variables $Z_{1000}$, et la comparer à la fonction de répartition de la gaussienne, qui est obtenue par la commande `scs.norm.cdf`.

(4) On tire au hasard $p \in [0,1]$ avec la commande `random.random()`, puis on observe des variables de Bernoulli $B_1, B_2, \ldots, B_n$ de paramètre $p$, qui reste inconnu. À l'aide du théorème central limite, construire un *estimateur* $\widehat{p}_n = \widehat{p}_n(B_1, \ldots, B_n)$ du paramètre $p$, puis un *intervalle de confiance* $[\widehat{p}_{n,-}, \widehat{p}_{n,+}]$ tel que

$$\lim_{n\to\infty} \mathbb{P}[p \in [\widehat{p}_{n,-}, \widehat{p}_{n,+}]] \geq 0.98.$$

Tester cet intervalle de confiance avec un échantillon de taille $n = 1000$.

La dernière question est la base théorique de la théorie des sondages.

### 3. Variables géométriques renormalisées.

Soit $\lambda > 0$ un paramètre fixé, et $X_n$ une variable aléatoire géométrique de paramètre $p = \frac{\lambda}{n}$ :

$$\mathbb{P}[X_n = k] = (1-p)^{k-1}\,p \text{ pour tout entier } k \geq 1.$$

On pose $Y_n = \frac{X_n}{n}$.

(1) Écrire un programme qui simule la variable $Y_n$. Dresser la fonction de répartition empirique de cette variable pour $n = 3, 10, 100$.

(2) Comparer le cas $n = 100$ avec la fonction de répartition théorique d'une variable réelle de loi exponentielle $\mathrm{Exp}(\lambda)$. Que peut-on conjecturer ?

(3) Que vaut $\mathbb{P}[Y_n \geq s]$ pour $s \in \mathbb{R}_+$ ? En déduire une preuve de la conjecture de la question précédente.

### 4. Maximum de variables aléatoires exponentielles.

Dans cet exercice, $X_1, X_2, \ldots$ est une suite de variables aléatoires indépendantes de loi $\mathrm{Exp}(1)$, et on s'intéresse à la suite des maxima :

$$L_n = \max(X_1, X_2, \ldots, X_n).$$

(1) Écrire un programme qui trace le graphe de $n \mapsto \frac{L_n}{\log n}$ pour $n \in [\![2, 5000]\!]$. A-t-on une convergence presque sûre ?

(2) Quelle est la fonction de répartition de $L_n$ ? Montrer que

$$\sum_{n\geq 1} \mathbb{P}[L_n \leq (1-\varepsilon)\log n] < +\infty$$

et que

$$\sum_{n\geq 1} \mathbb{P}[L_{\alpha_n} \geq (1+\varepsilon)\log \alpha_n] < +\infty$$

pour tout $\varepsilon > 0$, et pour une suite croissante $(\alpha_n)_{n\geq 1}$ appropriée. En déduire une preuve de la conjecture de la question précédente. On pourra s'appuyer sur le lemme de Borel-Cantelli pour contrôler

$$\liminf_{n\to\infty} \frac{L_n}{\log n} \quad \text{et} \quad \limsup_{n\to\infty} \frac{L_{\alpha_n}}{\log \alpha_n},$$

et pour la limite supérieure de $L_n$ (au lieu de $L_{\alpha_n}$), on pourra utiliser la croissance de la suite $(L_n)_{n\geq 1}$.

(3) Tracer maintenant le graphe de $n \mapsto L_n - \log n$ pour $n \in [\![1, 5000]\!]$. A-t-on une convergence presque sûre ?

(4) Calculer la limite de $\mathbb{P}[L_n \leq \log n + t]$, et en déduire la convergence en loi de la suite $(L_n - \log n)_{n\geq 1}$. La loi limite est appelée *loi de Gumbel*.

(5) Dessiner sur un même graphique la fonction de répartition empirique d'un échantillon de $N = 1000$ tirages de la variable $L_{1000} - \log 1000$, et la fonction de répartition de la loi de Gumbel.